A study on fuzzy logics for the formalization of gene expression data analysis and application to Bioinformatics

Joyanta Pal
Research Student
Department of Mathematics, Siksha Bhavan (Institute of Sciences)
Visva-Bharati (a central university)
Santiniketan- 731235, West Bengal
email- joyantamath@rediffmail.com

Supervisor: Professor Swapan Raha
Department of Mathematics, Siksha Bhavan (a Institute of Science)
Visva-Bharati (, a central university)
Santiniketan- 731235, West Bengal
email- swapan.raha@visva-bharati.ac.in

Registration No. - VB- 1691 OF 2012-13, Date:- 08.10.2012.
Date of Re-registration:- 08.10.2017.

Keywords:-
Fuzzy logic, Approximate reasoning, Gene expression, Bioinformatics

A Plan of work

Fuzzy logic deals with natural language description. It is considered to be the logic of inexact concepts. Linguistic modelling consistent with the qualitative linguistic description of biological systems [1], are inherently inexact. Accordingly, it is possible to express biological phenomenon in a typically natural way using fuzzy logic based techniques. Since it expresses natural events, it is found to be subjective in nature. Only a fuzzy set has the power and flexibility to represent such inexact natural expressions appropriately.

Zadeh observed that the term *fuzzy logic*[2] has two different meanings - wide and narrow.

> In a narrow sense, *fuzzy logic*, FLn, is a logical system which aims at a formalization of approximate reasoning. In this sense, FLn is an extension of multivalued logic. However, the agenda of FLn is quite different from that of traditional multivalued logics. In particular, such key concepts in FLn as the concept of a linguistic variable, canonical form, fuzzy if-then rule, fuzzy quantification and defuzzification, predicate modification, truth-qualification, the extension principle, the compositional rule of inference and interpolative reasoning, among others, are not addressed in traditional systems. This is the reason why FLn has a much wider range of applications than traditional systems.

> In its wide sense, *fuzzy logic*, FLw, is fuzzily synonymous with fuzzy set theory [3], FST, which is the theory of classes with unsharp boundaries. FST is much broader than FLn and includes the latter as one of its branches.

Recently, Petr Hajěk concluded that fuzzy logic is a logic — in its wide sense. Anything dealing with fuzziness may be called fuzzy logic. Whereas, in its narrow sense, it is a beautiful logic and important for applications. There are various systems of fuzzy logics. Formal calculi of non-traditional multivalued logic is the kernel or base of fuzzy logic.

Formally, fuzzy logic can be defined as an algebraic system $\langle L, \wedge, \vee \rangle$ where L, the truth set, is a complete residuated lattice. $\wedge$ is completely distributive with respect to $\vee$, $\wedge$ and $\vee$ are the well-known min and max operators. For convenience, L be the unit interval [0,1] where 0 is the bottom element and 1 is the top element of the lattice.

Many problems in bioinformatics have high dimensionality without sufficient number of corresponding examples (data points) that would allow the application of

well-known statistical techniques. Thus, results inferred from these data become statistically insignificant. This is why, fuzzy logic is used to handle imprecise or incomplete information extracted from huge data.

This research attempts to demonstrate the applicability of different fuzzy logics in problems of bioinformatics. Accordingly, we propose to study both type-1 and type-2 fuzzy logics for logical forms, their interpretations and derivations. It is already established that different fuzzy logics offer an appealing tool to model phenomenon described by a set of general knowledge of the form

if condition then prediction

best known as a fuzzy rule. A set of related fuzzy rules forms a fuzzy rule base that can be used to infer fuzzy values as results or prediction or decisions. These rules, express knowledge in the same way a human expert does.

It is based primarily on biological data which are typically large, have inherent biological complexity. Extracting knowledge from such data is a challenging task. This challenge requires sophisticated non-conventional tools for modelling. Fuzzy logic is a case in point. It is the flexibility, potentiality of fuzzy rules that allowed researchers to make an attempt to model considerable gene expressions data naturally.

Gene sequences [5] code proteins and maintain a complex process to regulate themselves. Different approaches have already been made over the years to analyze data for gene expression. In this research, an attempt is made to understand gene expression with the help of fuzzy set theory. Fuzzy set is considered to tackle the imprecision inherent in the natural language representation of the conditions for gene expression [6]. Recently, fuzzy rules were constructed to study the phenomenon of gene expression.

Genes work by making proteins. The qualitative aspect of gene expression [7], [8] is mainly due to production of the working proteins. Basal level synthesis always happens in the system. This means that proteins are always generated at a minimum level. The regulation of gene expression is a vital factor. This study deals formally with the manipulation of gene expression. Herein lies the motivation of this study.

A typical gene expression procedure could be represented using some imprecise yet natural statements. From this, a standard form can be constructed using rule structure — antecedent-consequent condition consisting of possibly imprecise concepts. This is necessary in order to capture the existing relationship among gene sequences in a natural way. Each concept can be represented by a set and operations, permissible under the set-theory could be used to manipulate them. Fuzzy set theory is a flexible, yet powerful, tool to model the uncertainty arising out of imprecise knowledge.

As an application to Bioinformatics, this research is expected to unravel among others, the functional and ancestral relationships between proteins using fuzzy alignment method.

References

[1] Andersson S., Eriksson K., *Dynamics of gene order structures and genomic architectures*, Kluwer Academic Publishers, 2009.

[2] Baldwin J.F., 'A new approach to approximate reasoning using a fuzzy logic', **Fuzzy sets and systems**, vol. 2, no. 3, 1979, 309-325.

[3] Dubois D., H. Prade, 'Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions', **Fuzzy sets and systems**, vol. 40, 1991, pp.143-202.

[4] Dubois D., H. Prade, 'Fuzzy sets in approximate reasoning, Part 2: Logical approaches', **Fuzzy sets and systems**, vol. 40, 1991, pp.203-244.

[5] Kari L., "Dna computing: arrival of biological mathematics," *The Mathematical Intelligencer*, 1997, vol. 19(2), pp. 9–22.

[6] Chen X., Kwong S., Li M., "A compression algorithm for dna sequences and its applications in genome comparison," in *Proceedings of the Sixth Annual International Computing and Combinatorics Conference-RECOMB, ACM Press*, 2000, pp. 107–117.

[7] Claverie J.M., "From bioinformatics to computational biology," *Genome Research*, 2000, vol. 10, pp. 1277–1279.

[8] Asyali M.H., Colak D., Demirkaya O., Inan M.S., "Gene expression profile classification: a review," *Current Bioinformatics*, 2006 vol. 1(1), pp. 55–73.